Luts H.[1], Jansen S. [1], Dreschler W. [2], & Wouters J. [1]

[1] ExpORL, Dept. Neurosciences, KU Leuven, Belgium
[2] Clinical and Experimental Audiology, Academic Medical Center Amsterdam, The Netherlands

Methods

Speech material

The Flemish/Dutch Matrix consists of 10 names, 10 verbs, 10 numerals, 10 colors and 10 objects (see Table 1). Within a column, all words have the same number of syllables. The verbs are all in the present tense, and the adjectives are all colors. The occurrence of phonemes in the base matrix closely resembles the reference distribution of phonemes in the Dutch language (Luyckx et al., 2007, Figure 1). A professional Belgian female speaker was selected. The speech material was recorded in a soundproof booth at the KU Leuven, Dept. Of Neurosciences, ExpORL, with an Edirol R-4 PRO recorder and a Sennheiser HS2 headset microphone, at a 44.1 kHz sampling rate and 24 bits resolution. A set of one hundred different sentences was recorded several times. In this set, each word occurred 10 times, always with the coarticulation to another adjacent word. The best recording of each sentence was selected, and these 100 sentences were equalized in rms (based on the average root-mean-square of a complete sentence, with silence parts included). The sentences were cut into words, preserving the coarticulation to the next word. This resulted in 500 different *.wav-files (each containing a word with the coarticulation to the following word). Subsequently, 50 lists of 10 sentences were generated with Matlab. Each list contained all words of the base matrix and was thus phonetically balanced. In this set of 500 sentences, each combination of a word with the coarticulation to the following word occurred exactly 5 times. All 500 sentences were included in the optimization measurements.

Table 1. The closed set of 50 words of the Flemish/Dutch Matrix

	Name	*Verb*	*Numeral*	*Color*	*Object*
1	Jeroen	heeft	twee	witte	fietsen
2	Thomas	kiest	drie	gele	manden
3	Lucas	koopt	vier	bruine	doeken
4	Jacob	zoekt	vijf	rode	dozen
5	Sofie	draagt	zes	blauwe	kousen
6	Ellen	maakt	acht	groene	bedden
7	Johan	wint	tien	grijze	jassen
8	Sara	krijgt	elf	zwarte	pennen
9	Emma	ziet	twaalf	paarse	ringen
10	David	leent	veel	beige	boten

Noise material

To generate the stationary speech-weighted noise the long-term average speech spectrum (LTASS) of the 500 sentences was determined. For each sentence, silence parts were removed (frames of 20 ms with RMS < 0.001) and the spectrum was calculated with a 4096-points fast Fourier transform using a rectangular window and without overlap. These spectra were then averaged, applying a weight according to the length of each sentence. For this LTASS, a 2048-taps finite impulse response filter was generated and applied on an 11 seconds long white noise. Transients at the start and the end of the speech-shaped noise were removed to allow looping of the noise without any clicks. The average RMS level of the noise was -27.0 dB FS.

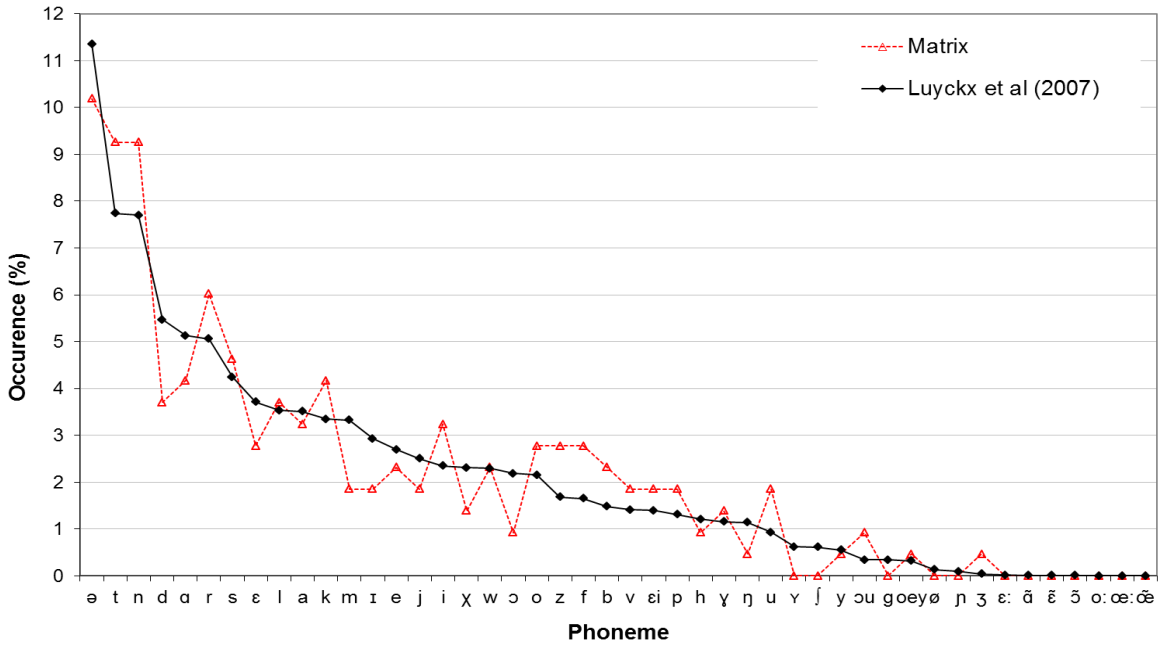

Figure 1. Phonetic distribution of the Flemish/Dutch Matrix (open triangles), compared to the mean phonetic distribution of spoken Dutch (Luyckx et al, 2007, filled diamonds)

Subjects

For the development of the Flemish/Dutch Matrix test, 98 normal-hearing Flemish subjects were tested. Hearing thresholds for the test ear were equal to or better than 20 dBHL for all octave frequencies between 250 and 8000 Hz. They were all native Dutch speakers from Belgium. Fifty-two normal-hearing adults (15 men) aged between 18 and 54 years (median age 22 years) participated in the optimization measures. Twenty-six normal-hearing adults (4 men) between 18 and 34 years old (mean age 21 years) participated in the selection phase. Twenty normal-hearing adults (7 men) between 18 and 25 years old (mean age 21 years) participated in the evaluation measurements.

Test set-up

All perceptual measurements were performed using a PC running Apex 3 software (Francart et al, 2008), a high-quality 24-bit RME sound card, and Sennheiser HDA200 headphones. The speech and noise were always presented monaurally to the subject's best ear. The setup was calibrated with a B&K sound level meter 2250 and a B&K artificial ear 4153. The noise started 500 ms before and ended 500 ms after each sentence, and was a randomly selected segment from the noise file. The noise level was always 65 dBSPL. The subjects received the base matrix on paper and were instructed to repeat the sentences as accurate and complete as possible. They were not obliged to guess if they were not sure. A word scoring procedure was used.

Optimization procedure

In order to reach the steepest possible slope at the SRT of the final reference psychometric curve, the speech material of the Flemish/Dutch Matrix was further optimized according to the procedure described by Wagener et al (2003) and Jansen et al (2012). Each subject listened to all 500 sentences (20 lists of 25 sentences) in two separate test sessions that started with one additional training list. The sentences were presented at 13 fixed SNRs (ranging from -20 to +4 dB in steps of 2 dB) and a noise level of 65 dBSPL. The SRT (speech reception threshold) of each of the 500 single words was then determined by applying a logistic regression fit to the data of all listeners together. To improve the homogeneity of the words with regard to their intelligibility, the level of each word was now adjusted towards the mean SRT. The level of the individual words was adjusted by maximally 6 dB.

This is more than in other languages where the maximal adjustment was typically 3 to 4 dB (e.g. Wagener et al, 2003; Jansen et al, 2012). However, the perceptual differences between 4 and 6 dB maximal adjustment were very limited. The 500 sentences were then regenerated and recombined into the original 50 lists of 10 sentences.

Selection of sentences

After the optimization procedure, the best test lists of 10 sentences were selected based on speech recognition scores and evaluation of naturalness. To obtain speech recognition scores, the optimized material was presented in double-lists (i.e. a combination of two original lists of 10 sentences that contained each word of the base matrix into one list of 20 sentences) to 20 normal-hearing listeners. The lists were presented at 4 fixed SNRs: -11.5, -10, -8.5 and -7 dB. The noise was kept fixed at 65 dBSPL. The SRT and slope were calculated for each original list of 10 sentences based on a logistic regression fit to the data of all listeners together. In addition, six normal-hearing listeners (4 students and 2 graduates in speech language pathology and audiology) evaluated all sentences with regard to the naturalness. The sentences were presented in quiet at a level of 55 dBSPL, and the naturalness of each sentence was rated on a 4-point scale: poor (0), moderate (1), good (2) or very good (3).

The total number of sentences was reduced from 500 to 260. In total 24 lists of 10 sentences were excluded: four lists showed a shallow list-specific slope (below 12%/dB), one list had a list-specific SRT that deviated more than 0.5 dB from the mean and 19 lists contained one or more unnaturally sounding sentences. The final speech material consisted of 26 balanced test lists of 10 sentences that contain each word of the matrix.

Evaluation procedure

To establish norms for normal-hearing listeners for the final speech material of the Flemish/Dutch Matrix, another group of twenty normal-hearing adults was tested. First, six double-lists were presented using the adaptive procedure of Brand & Kollmeier (2002). This allowed the evaluation of training effects and test-retest reliability. The noise was kept fixed at 65 dBSPL and the procedure started at an SNR of -4 dB. Subsequently, all 13 double-lists were presented at 4 fixed SNRs (-10.5, -9.0, -7.5 and -6.0 dB). The noise was again fixed at 65 dBSPL. By applying logistic regression fits to the data of each subject, the reference psychometric curve for normal-hearing listeners was determined.

Results

Based on the evaluation measurements with fixed SNR, the psychometric curve could be accurately determined for each subject for the final speech material. The average SRT was -9.5 dBSNR (with a standard deviation across subjects of 0.8 dB). The average slope was 13.9 %/dB (with a standard deviation of 1.5%/dB). Individual results for the 20 subjects can be found in Table 2 and Figure 2. Before optimization, the average subject-specific slope was only 8.7%/dB, indicating that the homogenization of the individual words resulted in a significant improvement of the steepness of the psychometric function.

Table 2. Speech recognitions scores per SNR and estimated SRT and slope for 20 normal-hearing listeners.

	Speech recognition score (%)				*SRT*	*Slope*
Subject	*-10,5 dB*	*-9 dB*	*-7,5 dB*	*-6 dB*	*(dBSNR)*	*(%/dB)*
1	50	61	81	91	-10,3	12,2
2	23	45	56	77	-8,3	12,2
3	27	52	72	83	-9,0	14,8
4	28	57	69	75	-9,0	11,9
5	19	44	62	84	-8,5	16,1
6	28	54	71	85	-9,1	14,8
7	44	66	78	86	-10,1	12,1
8	33	50	73	88	-9,2	14,7
9	28	61	69	83	-9,2	13,2
10	22	47	69	86	-8,7	17,1
11	34	54	76	86	-9,3	14,1
12	41	59	74	87	-9,7	12,0
13	45	64	78	88	-10,1	12,3
14	24	49	62	87	-8,7	15,2
15	57	75	87	92	-11,0	13,1
16	43	70	81	90	-10,2	14,5
17	47	68	84	94	-10,3	14,9
18	30	46	67	86	-8,8	14,4
19	47	70	86	91	-10,3	15,1
20	44	70	83	88	-10,3	14,1
average					**-9,5**	**13,9**
stdev					**0,8**	**1,5**

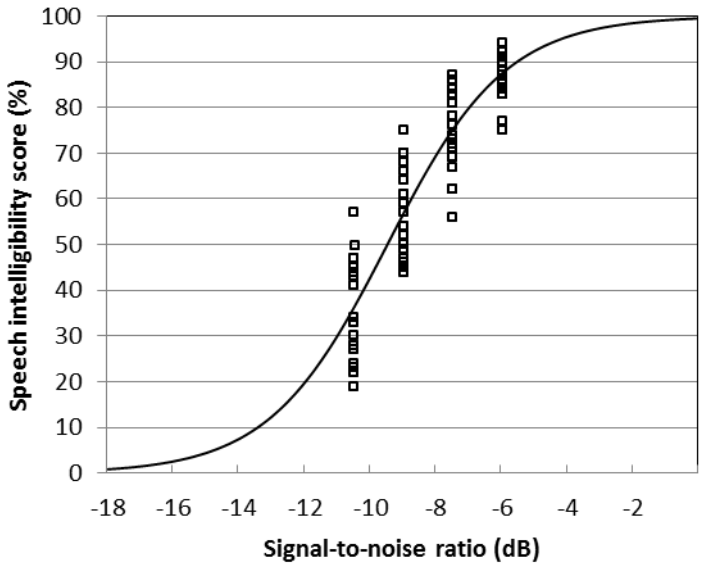

Figure 2. Reference psychometric function for normal-hearing listeners for the Flemish/Dutch Matrix test using word scoring. Squares represent subject scores per SNR.

List-specific SRTs were determined as well by pooling data of all subjects and applying the logistic regression fit per list. The standard deviation of the SRTs across the 13 double-lists was 0.2 dB, with a maximal deviation of an individual double-list from the overall mean SRT of only 0.4 dB. This is without correction for the performance level of the individual subjects. After correction, the standard deviation across lists remained 0.2 dB and the maximal deviation decreased to 0.3 dB (see Figure 3).

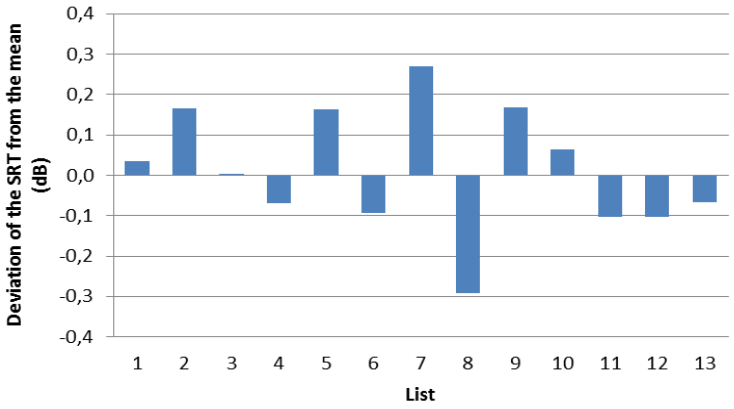

Figure 3. Deviation of the list-specific SRTs from the mean SRT for the 13 double-lists of the Flemish/Dutch Matrix, after correction for the performance level of individual subjects.

Figure 4 shows the results of the first six lists with an adaptive procedure. The training effect is, as expected, the largest from the first to the second list (0.7 dB). A repeated-measures ANOVA showed a significant training effect. Pairwise comparisons with Bonferroni correction showed that the first list differed significantly from all the following lists (always $p<0.008$) and the second list differed significantly from the fourth, fifth and sixth list (always $p<0.05$). From the third list on, the decrease in SRT was not significant. The test-retest reliability, defined as the root mean square of the within-subjects standard deviations of repeatedly measured adaptive SRTs, was calculated taking into account the SRTs of the third to the sixth list to exclude training effects. This resulted in a within-subject variability of 0.5 dB.

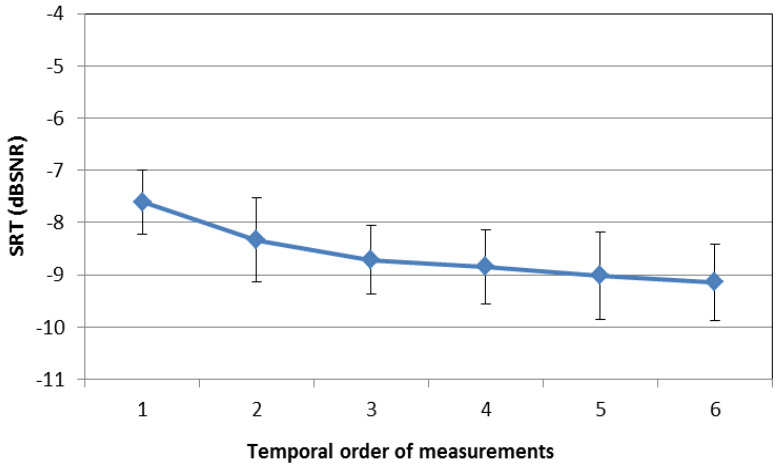

Figure 4. Training effect of the Flemish/Dutch Matrix test. For each measurement, one double-list of 20 sentences was used with an adaptive procedure.

Acknowledgements

The authors would like to thank Kelly Hendrickx, Annelie Slaets, Synthia Boon, Elisa Henry de Generet, Marieke Remmery and Ann-Sofie De Prins for their help in cutting the material and collecting data.

References

Brand T. & Kollmeier B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J. Acoust. Soc. Am.,* 111(6), 2801-2810.

Jansen S., Luts H., Wagener K.C., Kollmeier B., Del Rio M., Dauman R., Fraysse B., Vormès E., Frachet B., Wouters J., & van Wieringen A. (2012). Comparison of three types of French speech-in-noise tests: a multi-center study. *Int. J. Audiol.,* 51, 164-173.

Luyckx K., Kloots H., Coussé E. & Gillis S. (2007). Klankfrequenties in het Nederlands. In D. Sandra, R. Rymenans, P. Cuvelier, & P. Van Petegem, eds. *Tussen taal, spelling en onderwijs*. Ghent: Academia Press, pp. 141-154.

Wagener K.C., Josvassen J.L., & Ardenkjar R. (2003). Design, optimization and evaluation of a Danish sentence test in noise. *Int. J. Audiol.,* 42, 10-17.